Lecture 11: Conjugate Priors and Bayesian Regression

STA702

Merlise Clyde
Duke University

Semi-Conjugate Priors in Linear Regression

- Regression Model (Sampling model)

$$\mathbf{Y} \mid \boldsymbol{\beta}, \phi \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \phi^{-1}\mathbf{I}_n)$$

- Semi-Conjugate Prior Independent Normal Gamma

$$\begin{aligned}\boldsymbol{\beta} &\sim \mathsf{N}(\mathbf{b}_0, \boldsymbol{\Phi}_0^{-1}) \\ \phi &\sim \textsf{Gamma}(\nu_0/2, \textsf{SS}_0/2)\end{aligned}$$

  - Conditional Normal for $\boldsymbol{\beta} \mid \phi, \mathbf{Y}$ and
  - Conditional Gamma $\phi \mid \mathbf{Y}, \boldsymbol{\beta}$
  - requires Gibbs sampling or other Metropolis-Hastings algorithms

Conjugate Priors in Linear Regression

- Regression Model (Sampling model)

$$\mathbf{Y} \mid \boldsymbol{\beta}, \phi \sim \textsf{N}(\mathbf{X}\boldsymbol{\beta}, \phi^{-1}\mathbf{I}_n)$$

- Conjugate Normal-Gamma Model: factor joint prior $p(\boldsymbol{\beta}, \phi) = p(\boldsymbol{\beta} \mid \phi)p(\phi)$

$$\boldsymbol{\beta} \mid \phi \sim \textsf{N}(\mathbf{b}_0, \phi^{-1}\boldsymbol{\Phi}_0^{-1}) \qquad p(\boldsymbol{\beta} \mid \phi) = \frac{|\phi\boldsymbol{\Phi}_0|^{1/2}}{(2\pi)^{p/2}} e^{\left\{-\frac{\phi}{2}(\boldsymbol{\beta}-\mathbf{b}_0)^T\boldsymbol{\Phi}_0(\boldsymbol{\beta}-\mathbf{b}_0)\right\}}$$

$$\phi \sim \textsf{Gamma}(v_0/2, \textsf{SS}_0/2) \qquad p(\phi) = \frac{1}{\Gamma(\nu_0/2)}\left(\frac{\textsf{SS}_0}{2}\right)^{\nu_0/2}\phi^{\nu_0/2-1}e^{-}$$

$$\Rightarrow (\boldsymbol{\beta}, \phi) \sim \textsf{NG}(\mathbf{b}_0, \boldsymbol{\Phi}_0, \nu_o, \textsf{SS}_0)$$

- Normal-Gamma distribution indexed by 4 hyperparameters
- Note Prior Covariance for $\boldsymbol{\beta}$ is scaled by $\sigma^2 = 1/\phi$

Finding the Posterior Distribution

- Likelihood: $\mathcal{L}(\beta, \phi) \propto \phi^{n/2} e^{-\frac{\phi}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}$

$$p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n}{2}} e^{-\frac{\phi}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})} \times \phi^{\frac{\nu_0}{2}-1} e^{-\phi\frac{\textsf{SS}_0}{2}} \times \phi^{\frac{p}{2}} e^{-\frac{\phi}{2}(\boldsymbol{\beta}-\mathbf{b}_0)^T \boldsymbol{\Phi}_0(\boldsymbol{\beta}-\mathbf{b}_0)}$$

- Quadratic in Exponential

$$\exp\left\{-\frac{\phi}{2}(\boldsymbol{\beta}-\mathbf{b})^T \boldsymbol{\Phi}(\boldsymbol{\beta}-\mathbf{b})\right\} = \exp\left\{-\frac{\phi}{2}(\boldsymbol{\beta}^T \boldsymbol{\Phi}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\boldsymbol{\Phi}\mathbf{b} + \mathbf{b}^T\boldsymbol{\Phi}\mathbf{b})\right\}$$

  - Expand quadratics and regroup terms
  - Read off posterior precision from Quadratic in $\boldsymbol{\beta}$
  - Read off posterior mean from Linear term in $\boldsymbol{\beta}$
  - will need to complete the quadratic in the posterior mean due to ϕ

Expand and Regroup

$$p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n}{2}} e^{-\frac{\phi}{2}(\mathbf{Y}^T\mathbf{Y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta})} \times$$
$$\phi^{\frac{\nu_0}{2}-1} e^{-\phi\frac{\mathsf{SS}_0}{2}} \times \phi^{\frac{p}{2}} e^{-\frac{\phi}{2}(\boldsymbol{\beta}\boldsymbol{\Phi}_0\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\boldsymbol{\Phi}_0\mathbf{b} + \mathbf{b}_0^T\boldsymbol{\Phi}_0\mathbf{b}_0)}$$

$$p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\mathsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T\boldsymbol{\Phi}_0\mathbf{b}_0)} \times$$
$$e^{-\frac{\phi}{2}(\boldsymbol{\beta}^T(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} + \boldsymbol{\beta}\boldsymbol{\Phi}_0\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\boldsymbol{\Phi}_0\mathbf{b})}$$

$$= \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\mathsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T\boldsymbol{\Phi}_0\mathbf{b}_0)} \times$$
$$e^{-\frac{\phi}{2}(\boldsymbol{\beta}^T(\mathbf{X}^T\mathbf{X} + \boldsymbol{\Phi}_0)\boldsymbol{\beta})} \times$$
$$e^{-\frac{\phi}{2}\left(-2\boldsymbol{\beta}^T(\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\Phi}_0\mathbf{b}_0)\right)}$$

Complete the Quadratic

$$p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\mathsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T \boldsymbol{\Phi}_0 \mathbf{b}_0)} \times$$
$$e^{-\frac{\phi}{2}(\boldsymbol{\beta}^T(\mathbf{X}^T\mathbf{X} + \boldsymbol{\Phi}_0)\boldsymbol{\beta})} \times$$
$$e^{-\frac{\phi}{2}\left(-2\boldsymbol{\beta}^T \boldsymbol{\Phi}_n \boldsymbol{\Phi}_n^{-1}(\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\Phi}_0\mathbf{b}_0)\right)} \times$$
$$e^{-\frac{\phi}{2}(\mathbf{b}_n^T\boldsymbol{\Phi}_n\mathbf{b}_n - \mathbf{b}_n^T\boldsymbol{\Phi}_n\mathbf{b}_n)}$$

$$\boldsymbol{\Phi}_n \equiv \mathbf{X}^T\mathbf{X} + \boldsymbol{\Phi}_0$$

$$\mathbf{b}_n \equiv \boldsymbol{\Phi}_n^{-1}(\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\Phi}_0\mathbf{b}_0)$$

$$= \phi^{\frac{n+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\mathsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T\boldsymbol{\Phi}_0\mathbf{b}_0 - \mathbf{b}_n^T\boldsymbol{\Phi}_n\mathbf{b}_n)} \times$$
$$\phi^{\frac{p}{2}} e^{-\frac{\phi}{2}\left((\boldsymbol{\beta}^T - \mathbf{b}_n)^T\boldsymbol{\Phi}_n(\boldsymbol{\beta} - \mathbf{b}_n)\right)}$$

$$\propto \phi^{\frac{n+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\mathsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T\boldsymbol{\Phi}_0\mathbf{b}_0 - \mathbf{b}_n^T\boldsymbol{\Phi}_n\mathbf{b}_n)} \times$$
$$|\phi\boldsymbol{\Phi}_n|^{\frac{1}{2}} e^{-\frac{\phi}{2}\left((\boldsymbol{\beta}^T - \mathbf{b}_n)^T\boldsymbol{\Phi}_n(\boldsymbol{\beta} - \mathbf{b}_n)\right)}$$

Posterior Distributions

Posterior density (up to normalizing contants) $p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) = p(\phi \mid \mathbf{Y}) p(\boldsymbol{\beta} \mid \phi \mathbf{Y})$

$$p(\phi \mid \mathbf{Y}) p(\boldsymbol{\beta} \mid \phi \mathbf{Y}) \propto \phi^{\frac{n+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\textsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T \boldsymbol{\Phi}_0 \mathbf{b}_0 - \mathbf{b}_n^T \boldsymbol{\Phi}_n \mathbf{b}_n)} \times \\ (2\pi)^{-\frac{p}{2}} |\phi \boldsymbol{\Phi}_n|^{\frac{1}{2}} e^{-\frac{\phi}{2}(\boldsymbol{\beta} - \mathbf{b}_n)^T \boldsymbol{\Phi}_n (\boldsymbol{\beta} - \mathbf{b}_n)}$$

Marginal

$$\begin{aligned} p(\phi \mid \mathbf{Y}) \propto & \ \phi^{\frac{n+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\textsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T \boldsymbol{\Phi}_0 \mathbf{b}_0 - \mathbf{b}_n^T \boldsymbol{\Phi}_n \mathbf{b}_n)} \times \\ & \int_{\mathbb{R}^p} (2\pi)^{-\frac{p}{2}} |\phi \boldsymbol{\Phi}_n|^{\frac{1}{2}} e^{-\frac{\phi}{2}(\boldsymbol{\beta} - \mathbf{b}_n)^T \boldsymbol{\Phi}_n (\boldsymbol{\beta} - \mathbf{b}_n)} \, d\boldsymbol{\beta} \\ = & \ \phi^{\frac{n+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\textsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T \boldsymbol{\Phi}_0 \mathbf{b}_0 - \mathbf{b}_n^T \boldsymbol{\Phi}_n \mathbf{b}_n)} \end{aligned}$$

- Conditional Normal for $\boldsymbol{\beta} \mid \phi, \mathbf{Y}$ and marginal Gamma for $\phi \mid \mathbf{Y}$
- No need for Gibbs sampling!

NG Posterior Distribution

$$\begin{aligned}\boldsymbol{\beta} \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{b}_n, (\phi\boldsymbol{\Phi}_n)^{-1}) \\ \phi \mid \mathbf{Y} &\sim \mathsf{Gamma}(\frac{\nu_n}{2}, \frac{\mathsf{SS}_n}{2}) \\ (\boldsymbol{\beta}, \phi) \mid \mathbf{Y} &\sim \mathsf{NG}(\mathbf{b}_n, \boldsymbol{\Phi}_n, \nu_n, \mathsf{SS}_n)\end{aligned}$$

Hyperparameters:

$$\begin{aligned}\boldsymbol{\Phi}_n &= \mathbf{X}^T\mathbf{X} + \boldsymbol{\Phi}_0 & \mathbf{b}_n &= \boldsymbol{\Phi}_n^{-1}(\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\Phi}_0\mathbf{b}_0) \\ \nu_n &= n + \nu_0 & \mathsf{SS}_n &= \mathsf{SS}_0 + \mathbf{Y}^T\mathbf{Y} + \mathbf{b}_0^T\boldsymbol{\Phi}_0\mathbf{b}_0 - \mathbf{b}_n^T\boldsymbol{\Phi}_n\mathbf{b}_n\end{aligned}$$

$$\begin{aligned}\mathsf{SS}_n &= \mathsf{SS}_0 + \|\mathbf{Y} - \mathbf{X}\mathbf{b}_n\|^2 + (\mathbf{b}_0 - \mathbf{b}_n)^T\boldsymbol{\Phi}_0(\mathbf{b}_0 - \mathbf{b}_n) \\ &= \mathsf{SS}_0 + \|\mathbf{Y} - \mathbf{X}\mathbf{b}_n\|^2 + \|\mathbf{b}_0 - \mathbf{b}_n\|^2_{\boldsymbol{\Phi}_0}\end{aligned}$$

- Inner product induced by prior precision $\langle u, v \rangle_A \equiv u^T A v$
- $\|\mathbf{b}_0 - \mathbf{b}_n\|^2_{\boldsymbol{\Phi}_0}$ mismatch of prior and posterior mean under prior

Marginal Distribution

▼ **Theorem:** Student-t

Let $\boldsymbol{\theta} \mid \phi \sim \mathsf{N}(m, \frac{1}{\phi}\boldsymbol{\Sigma})$ and $\phi \sim \mathsf{Gamma}(\nu/2, \nu\hat{\sigma}^2/2)$.

Then $\boldsymbol{\theta}\ (p \times 1)$ has a p dimensional multivariate t distribution

$$\boldsymbol{\theta} \sim t_\nu(m, \hat{\sigma}^2\boldsymbol{\Sigma})$$

with location m, scale matrix $\hat{\sigma}^2\boldsymbol{\Sigma}$ and density

$$p(\boldsymbol{\theta}) \propto \left[1 + \frac{1}{\nu}\frac{(\boldsymbol{\theta} - m)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} - m)}{\hat{\sigma}^2}\right]^{-\frac{p+\nu}{2}}$$

Note - true for prior or posterior given $\mathbf{Y}$

Derivation

Marginal density $p(\boldsymbol{\theta}) = \int_0^\infty p(\boldsymbol{\theta} \mid \phi) p(\phi) \, d\phi$

$$\begin{aligned}
p(\boldsymbol{\theta}) &\propto \int |\boldsymbol{\Sigma}/\phi|^{-1/2} e^{-\frac{\phi}{2}(\boldsymbol{\theta}-m)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-m)} \phi^{\nu/2-1} e^{-\phi\frac{\nu\hat{\sigma}^2}{2}} \, d\phi \\
&\propto \int \phi^{p/2} \phi^{\nu/2-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}} \, d\phi \\
&\propto \int \phi^{\frac{p+\nu}{2}-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}} \, d\phi \\
&= \Gamma((p+\nu)/2) \left(\frac{(\boldsymbol{\theta}-m)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2} \right)^{-\frac{p+\nu}{2}} \\
&\propto \left((\boldsymbol{\theta}-m)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2 \right)^{-\frac{p+\nu}{2}} \\
&\propto \left(1 + \frac{1}{\nu} \frac{(\boldsymbol{\theta}-m)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-m)}{\hat{\sigma}^2} \right)^{-\frac{p+\nu}{2}}
\end{aligned}$$

Marginal Posterior Distribution of β

$$\boldsymbol{\beta} \mid \phi, \mathbf{Y} \sim \mathsf{N}(\mathbf{b}_n, \phi^{-1}\boldsymbol{\Phi}_n^{-1})$$
$$\phi \mid \mathbf{Y} \sim \mathsf{Gamma}\left(\frac{\nu_n}{2}, \frac{\mathsf{SS}_n}{2}\right)$$

- Let $\hat{\sigma}^2 = \mathsf{SS}_n/\nu_n$ (Bayesian MSE)
- The marginal posterior distribution of $\boldsymbol{\beta}$ is multivariate Student-t

$$\boldsymbol{\beta} \mid \mathbf{Y} \sim t_{\nu_n}(\mathbf{b}_n, \hat{\sigma}^2\boldsymbol{\Phi}_n^{-1})$$

- Any linear combination $\lambda^T\boldsymbol{\beta}$ has a univariate t distribution with $\mathbf{v}_n$ degrees of freedom

$$\lambda^T\boldsymbol{\beta} \mid \mathbf{Y} \sim t_{\nu_n}(\lambda^T\mathbf{b}_n, \hat{\sigma}^2\lambda^T\boldsymbol{\Phi}_n^{-1}\lambda)$$

- use for individual $\boldsymbol{\beta}_j$, the mean of Y, $\mathbf{x}^T\boldsymbol{\beta}$, at $\mathbf{x}$, or predictions $Y^* = \mathbf{x}^{*T}\boldsymbol{\beta} + \epsilon_i^*$

Predictive Distributions

Suppose $\mathbf{Y}^* \mid \boldsymbol{\beta}, \phi \sim \mathsf{N}_s(\mathbf{X}^*\boldsymbol{\beta}, \mathbf{I}_s/\phi)$ and is conditionally independent of $\mathbf{Y}$ given $\boldsymbol{\beta}$ and ϕ

- What is the predictive distribution of $\mathbf{Y}^* \mid \mathbf{Y}$?
- Use the representation that $\mathbf{Y}^* \overset{\text{D}}{=} \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*$ and $\boldsymbol{\epsilon}^*$ is independent of $\mathbf{Y}$ given ϕ

$$\begin{aligned}
\mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^* \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*\boldsymbol{\Phi}_n^{-1}\mathbf{X}^{*T} + \mathbf{I}_s)/\phi) \\
\mathbf{Y}^* \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*\Phi_n^{-1}\mathbf{X}^{*T} + \mathbf{I}_s)/\phi) \\
\phi \mid \mathbf{Y} &\sim \textsf{Gamma}\left(\frac{\nu_n}{2}, \frac{\hat{\sigma}^2\nu_n}{2}\right)
\end{aligned}$$

- Use the Theorem to conclude that

$$\mathbf{Y}^* \mid \mathbf{Y} \sim t_{\nu_n}(\mathbf{X}^*\mathbf{b}_n, \hat{\sigma}^2(\mathbf{I} + \mathbf{X}^*\boldsymbol{\Phi}_n^{-1}\mathbf{X}^T))$$

Choice of Conjugate (or Semi-Conjugate) Prior

- need to specify Normal prior mean $\mathbf{b}_0$ and precision $\mathbf{\Phi}_0$
- need to specify Gamma shape (ν_o prior df) and rate (estimate of σ^2)
- hard in higher dimensions!
- default choices?
  - Jeffreys' prior
  - unit-information prior
  - Zellner's g-prior
  - ridge priors
  - mixtures of conjugate priors

Jeffreys' Prior

- Jeffreys prior is invariant to model parameterization of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi)$

$$p(\boldsymbol{\theta}) \propto |\mathcal{I}(\boldsymbol{\theta})|^{1/2}$$

- $\mathcal{I}(\boldsymbol{\theta})$ is the Expected Fisher Information matrix

$$\mathcal{I}(\theta) = -\mathsf{E}[\left[\frac{\partial^2 \log(\mathcal{L}(\boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j}\right]]$$

- log likelihood expressed as function of sufficient statistics

$$\log(\mathcal{L}(\boldsymbol{\beta}, \phi)) = \frac{n}{2}\log(\phi) - \frac{\phi}{2}\|(\mathbf{I}_n - \mathbf{P}_{\mathbf{x}})\mathbf{Y}\|^2 - \frac{\phi}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

- projection matrix $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$

Information matrix

$$\frac{\partial^2 \log \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{bmatrix} -\phi(\mathbf{X}^T\mathbf{X}) & -(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ -(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X}) & -\frac{n}{2}\frac{1}{\phi^2} \end{bmatrix}$$

$$\mathsf{E}[\frac{\partial^2 \log \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}] = \begin{bmatrix} -\phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & -\frac{n}{2}\frac{1}{\phi^2} \end{bmatrix}$$

$$\mathcal{I}((\boldsymbol{\beta}, \phi)^T) = \begin{bmatrix} \phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & \frac{n}{2}\frac{1}{\phi^2} \end{bmatrix}$$

Jeffreys' Prior (don't use!)

$$p_J(\boldsymbol{\beta}, \phi) \propto |\mathcal{I}((\boldsymbol{\beta}, \phi)^T)|^{1/2} = |\phi \mathbf{X}^T\mathbf{X}|^{1/2}\left(\frac{n}{2}\frac{1}{\phi^2}\right)^{1/2} \propto \phi^{p/2-1}|\mathbf{X}^T\mathbf{X}|^{1/2}$$
$$\propto \phi^{p/2-1}$$

Recommended Independent Jeffreys Prior

- Treat $\boldsymbol{\beta}$ and ϕ separately (*orthogonal parameterization*)
- $p_{IJ}(\boldsymbol{\beta}) \propto |\mathcal{I}(\boldsymbol{\beta})|^{1/2}$ and $p_{IJ}(\phi) \propto |\mathcal{I}(\phi)|^{1/2}$

$$\mathcal{I}((\boldsymbol{\beta}, \phi)^T) = \begin{bmatrix} \phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & \frac{n}{2}\frac{1}{\phi^2} \end{bmatrix}$$

$$\begin{aligned} p_{IJ}(\boldsymbol{\beta}) &\propto |\phi\mathbf{X}^T\mathbf{X}|^{1/2} \propto 1 \\ p_{IJ}(\phi) &\propto \phi^{-1} \\ p_{IJ}(\beta, \phi) &\propto p_{IJ}(\boldsymbol{\beta}) p_{IJ}(\phi) = \phi^{-1} \end{aligned}$$